# Automatic Age-Related Macular Degeneration Detection and Staging

Mark J. J. P. van Grinsven[a], Yara T. E. Lechanteur[b], Johannes P. H. van de Ven[b], Bram van Ginneken[a], Thomas Theelen[b] and Clara I. Sánchez[a]

[a]Radboud University Nijmegen Medical Centre, Diagnostic Image Analysis Group.
[b]Radboud University Nijmegen Medical Centre, Department of Ophthalmology.

## ABSTRACT

Age-related macular degeneration (AMD) is a degenerative disorder of the central part of the retina, which mainly affects older people and leads to permanent loss of vision in advanced stages of the disease. AMD grading of non-advanced AMD patients allows risk assessment for the development of advanced AMD and enables timely treatment of patients, to prevent vision loss. AMD grading is currently performed manually on color fundus images, which is time consuming and expensive. In this paper, we propose a supervised classification method to distinguish patients at high risk to develop advanced AMD from low risk patients and provide an exact AMD stage determination. The method is based on the analysis of the number and size of drusen on color fundus images, as drusen are the early characteristics of AMD. An automatic drusen detection algorithm is used to detect all drusen. A weighted histogram of the detected drusen is constructed to summarize the drusen extension and size and fed into a random forest classifier in order to separate low risk from high risk patients and to allow exact AMD stage determination. Experiments showed that the proposed method achieved similar performance as human observers in distinguishing low risk from high risk AMD patients, obtaining areas under the Receiver Operating Characteristic curve of 0.929 and 0.934. A weighted kappa agreement of 0.641 and 0.622 versus two observers were obtained for AMD stage evaluation. Our method allows for quick and reliable AMD staging at low costs.

**Keywords:** Age-related macular degeneration, Grading, Supervised classification

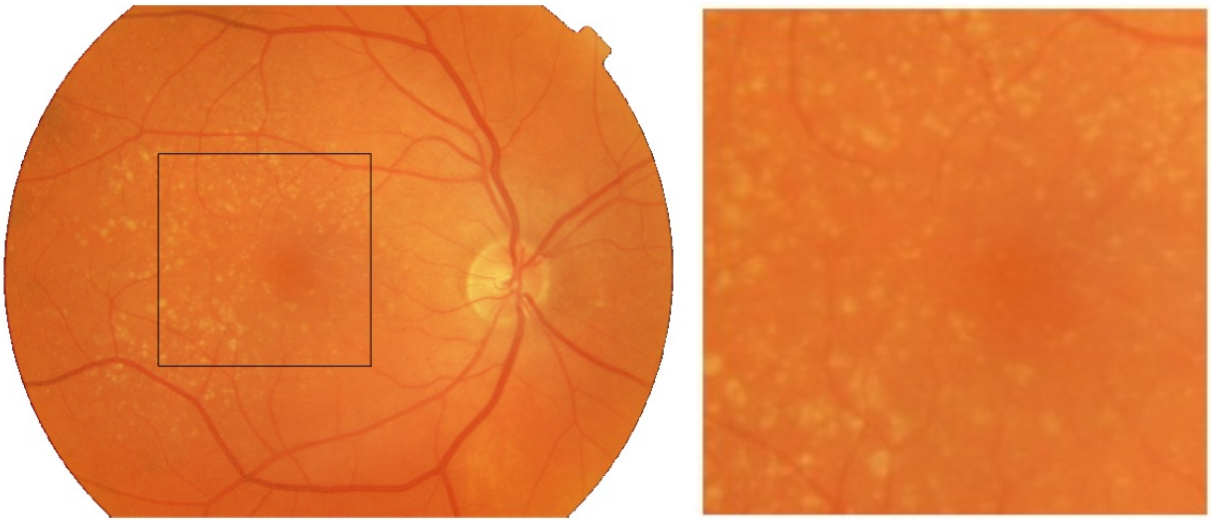
## 1. INTRODUCTION

Age-related Macular Degeneration (AMD) is the most important cause of legal blindness amongst the elder people in developed countries.[1] AMD can progress from early and intermediate stages, with no or subtle visual changes, to an advanced stage, where the loss of central vision can occur quickly. Recent studies showed that patients with intermediate AMD are at higher risk of developing advanced AMD.[2] Timely detection of those patients at risk is crucial to instruct patients and to apply prophylactic regimen such as vitamin supplementation to slow down the progression of the disease.[3,4] Therefore, identification of high risk patients is a key for saving vision.

Identification of patients at risk is currently performed using color fundus images by manually determining the size and extension of drusen, yellowish-white subretinal deposits that are the hallmark features of early and intermediate AMD, besides other retinal changes such as retinal pigmentary abnormalities.[5] Figure 1 shows an image containing drusen. However, human grading is time-consuming and prone to inter-observer variations.[6] Automating this process can significantly reduce the specialist's workload and pave the way for a cost-efficient screening program for risk patients. Several methods have been proposed in the literature to automatically determine the presence of drusen.[7,8] However, only the presence of drusen is not directly correlated with the risk of progression to advanced AMD. The presence of a few small drusen is normal with advancing age but it is not a sign of underlying AMD. Drusen number, location, area and size also have to be taken into account for this. Other methods only quantified the drusen extension without determining the disease stage or patients at high risk.[9–12] We are not aware of a software that automatically detects patients at high risk and differentiates between AMD stages based on drusen extension and size.

In this paper, we propose a method for the automatic identification of patients who are at high risk of developing advanced AMD and for exact AMD stage grading based on the analysis of drusen extension and size on color fundus images. Drusen extension and size is encoded by means of a weighted histogram using the outcome of an automatic drusen detection system developed in our group. A supervised classification method makes use of this information to detect images of high risk patients and to determine the disease stage.

Medical Imaging 2013: Computer-Aided Diagnosis, edited by Carol L. Novak, Stephen Aylward, Proc. of SPIE
Vol. 8670, 86700M · © 2013 SPIE · CCC code: 1605-7422/13/$18 · doi: 10.1117/12.2007563



(a) Color fundus image containing drusen (b) Zoomed in

Figure 1. (a): Example color fundus image containing drusen, yellowish-white spots in the image. (b) shows the zoomed in region indicated by the black box on (a).

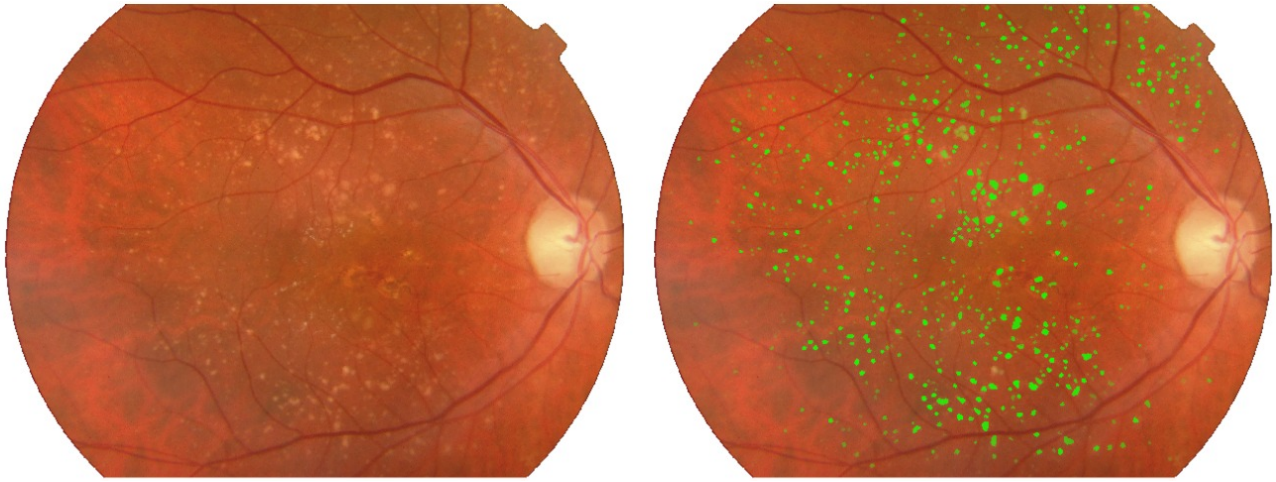
## 2. MATERIALS AND METHODS

### 2.1 Materials

For this study, 257 images of eyes of non-advanced AMD patients and control subjeccts were randomly selected from the European Genetic Database[13,14](EUGENDA), a large multicenter database for clinical and molecular analysis of AMD. A Topcon TRC 501X digital fundus camera at 50° (Topcon Corporation, Tokyo, Japan) and a Canon CR-DGi non-mydriatic retinal camera at 45° (Canon Inc., Tokyo, Japan) were used to acquire color fundus images centered on the macular region. Before imaging, pupil dilation was achieved with topical 1.0% tropicamide and 2.5% phenylephrine. Images from both cameras were merged into one dataset. The images were graded independently by two trained graders (Observer 1 and Observer 2) using the staging criteria adopted from the Wisconsin age-related maculopathy grading system, classifying images into no AMD, early AMD and intermediate AMD.[5] According to Observer 1, 118 images were identified as no AMD, 64 images as early AMD and 75 images as intermediate AMD; whereas Observer 2 annotated 120 images with no AMD, 61 images with early AMD and 76 images with intermediate AMD.

### 2.2 Automatic drusen detection system

In order to determine the drusen extension and size, a method to automatically detect drusen in color fundus images is applied. The method is based on three different steps: a candidate extraction step, a segmentation step and a drusen classification step. In the first step, a pixel probability map representing the pixel probability to belong to a bright lesion is obtained, using a k-nearest neighbor (kNN) classifier and a group of features extracted from Gaussian filter banks.[15] In the second step, drusen candidate lesions are segmented using dynamic programming with the local maxima from the pixel probability map as seed points. After this step, a group of candidate lesions $\{1, ..., i, ..., N\}$ are segmented in the image. In the last step, the candidate lesions are assigned a probability of being a true druse using a Linear Discriminant classifier (LDA). For this last step, a large set of features including color, shape, context and intensity features are used. The classifiers used in these steps are trained with a separate group of training images, which are not used for our present study.[15] The output of the drusen detection system is a drusen probability map where $p_i$ represents the probability of the segmented lesion $i$ of being a true druse. See Figure 2 for an example of the output of the drusen detection system.



(a) Color fundus image (b) Drusen detected on the color fundus image

Figure 2. (a): Color fundus image. (b): Automatic detected drusen on (a). Brighter color represents a higher probability of being a true druse.

### 2.3 Analysis of drusen extension and size

To encode the drusen extension and size of an image, a weighted histogram is created. Let $d_i$ be the size ($\mu m$) of the segmented lesion $i$, calculated as its largest diameter. Given $p_i$ and $d_i$, the value $h_n$ of the histogram bin $n$ is defined as:

$$h_n = \sum_{i \in L_n} p_i \tag{1}$$

where $L_n$ is the group of lesions whose size is $\tau n \leq d_i < \tau(n+1)$. $\tau$ and $n$ control the bin size and the histogram resolution, respectively. For this study, $n = 0, ..., 36$ and $\tau = 10\mu m$ was used. Note that the last bin ($n = 36$) takes into account all the drusen with sizes $d_i$ larger than 360 $\mu m$. Figure 3 shows some example images with different AMD stages and the corresponding histograms.

### 2.4 Identification of high-risk patients and AMD staging

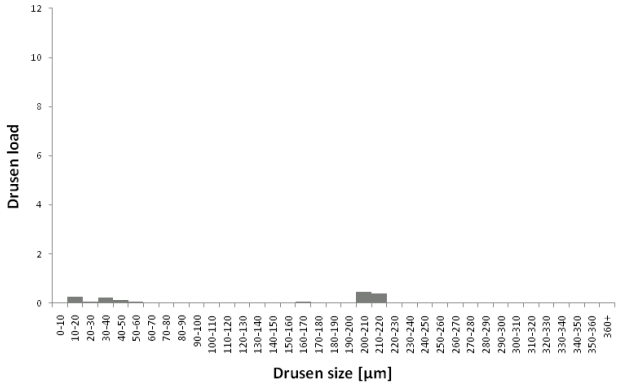
Using the previously described histogram as features, a supervised classifier is trained to separate images of high risk patients (with intermediate AMD) from images of low risk patients (no AMD or early AMD). In a pilot experiment, different classifiers were tested for this task, namely a Linear Discriminant classifier (LDA), a k-nearest neighbor classifier (kNN), a gentle boost classifier (GBC), a random forest classifier (RF) and a support vector machine classifier (SVM). The RF classifier slightly outperformed the other classifiers. Therefore, only the results of the experiments using the RF are reported.

For determining the exact AMD stage, a one-versus-one voting scheme[16] using RF classifiers and the histogram features is used to classify the image into no AMD, early AMD or intermediate AMD. If the voting tied, the image was assigned to the stage early AMD because this stage is the most ambiguous stage, being in between the other two stages.

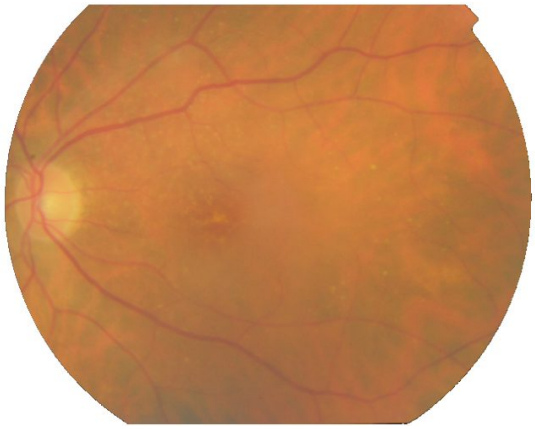
## 3. RESULTS

Receiver Operating Characteristics (ROC) analysis was performed using a leave-one-out procedure to evaluate the system performance in identifying high risk patients. Due to the lack of a single gold-standard, the system was evaluated first taking the annotations of Observer 1 as reference standard and then taking the annotations of Observer 2 as reference standard. Figure 4 shows the corresponding ROC curves. Areas under the ROC curve of 0.929 and 0.934 were obtained using Observer 1 and 2 as reference standard, respectively. For the cut-off point in the ROCs which maximizes sensitivity + specificity, the CAD system reached sensitivity/specificity pairs of 0.885/0.867 and 0.895/0.882 using Observer 1 and Observer 2 as reference, respectively. Observer 1 has a sensitivity/specificity pair of 0.842/0.939, whereas Observer 2 performs at 0.853/0.934, when using the other observer as reference.

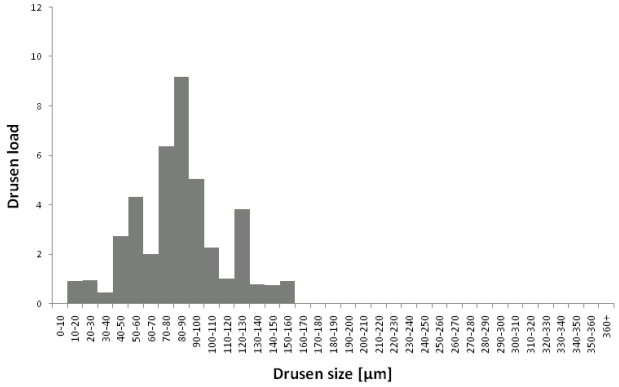
(a) Sample color fundus image without AMD




(b) No AMD: Histogram of the image shown in (a) representing small drusen load

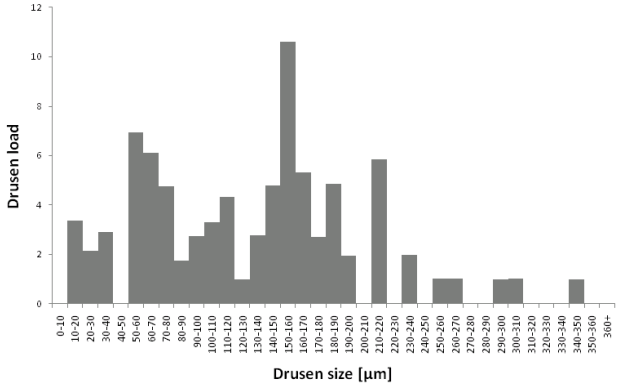


(c) Sample color fundus image of early AMD



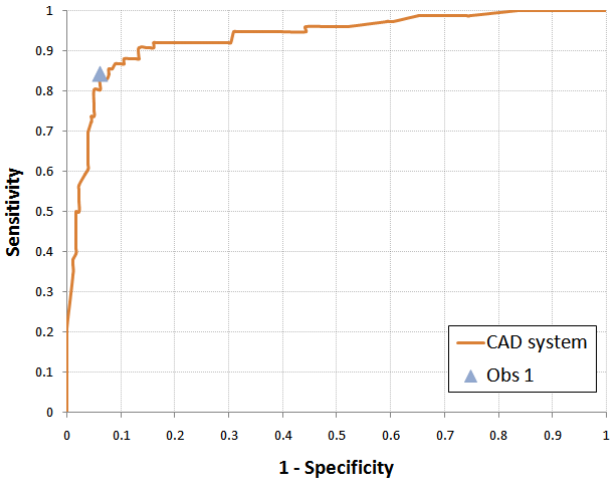

(d) Early AMD: Histogram of the image shown in (c) representing intermediate drusen load

(e) Sample color fundus image of intermediate AMD

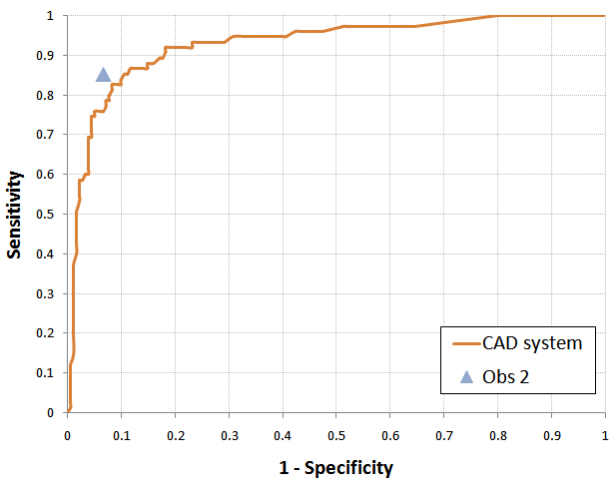



(f) Intermediate AMD: Histogram of the image shown in (e) representing large drusen load

Figure 3. (a),(c),(e): Color fundus image in different AMD stage. (b),(d),(f): Weighted histogram of images (a),(c) and (e), respectively.




(a) ROC 1




(b) ROC 2

Figure 4. ROC curves of the proposed system for distinguishing high risk patients (with intermediate AMD) from low risk patients (with no AMD or early AMD). (a) Annotations of Observer 2 are used as reference standard. The operating point of Observer 1 (Obs 1) is plotted as well. CAD system reached an area under the ROC of 0.934. (b): Annotations of Observer 1 are used as reference standard. The performance of Observer 2 (Obs 2) is plotted as well. CAD system reached an area under the ROC of 0.929.

To evaluate the system performance in determining the exact AMD stage, the kappa ($\kappa$) agreement between the proposed system and both observers was calculated using a leave-one-out scheme. The confusion matrices and the corresponding weighted kappa agreement for the automatic system and the observers are shown in Table 1.

Table 1. Confusion matrix of staging AMD into 1: no AMD, 2: early AMD and 3: intermediate AMD. (a) Results of the automatic system (CAD) vs observer 1 (Obs 1), (b) CAD vs and observer 2 (Obs 2) and (c) Obs 1 vs Obs 2. Kappa ($\kappa$) agreement and 95% confidence interval (CI) are also reported in the tables.

(a) CAD vs Obs 1

| | | CAD | | |
|---|---|---|---|---|
| | stage | 1 | 2 | 3 |
| | 1 | 99 | 15 | 4 |
| Obs 1 | 2 | 36 | 23 | 5 |
| | 3 | 6 | 8 | 61 |
| | $\kappa$=0.641 | | | |
| | 95% CI = 0.563-0.718 | | | |

(b) CAD vs Obs 2

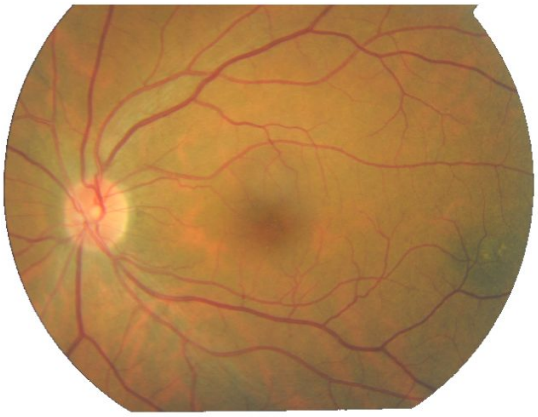
| | | CAD | | |
|---|---|---|---|---|
| | stage | 1 | 2 | 3 |
| | 1 | 98 | 17 | 5 |
| Obs 2 | 2 | 31 | 21 | 9 |
| | 3 | 6 | 10 | 60 |
| | $\kappa$=0.622 | | | |
| | 95% CI = 0.544-0.700 | | | |

(c) Obs1 vs Obs 2

| | | Obs 1 | | |
|---|---|---|---|---|
| | stage | 1 | 2 | 3 |
| | 1 | 103 | 15 | 2 |
| Obs 2 | 2 | 13 | 39 | 9 |
| | 3 | 2 | 10 | 64 |
| | $\kappa$=0.765 | | | |
| | 95% CI = 0.704-0.827 | | | |

## 4. DISCUSSION

In this paper a supervised classification method automatic for identification of patients who are at high risk of developing advanced AMD as well as for exact AMD stage classification was presented. Our system operates at performance equal to human observers for the identification of high risk patients and shows promising results for exact AMD stage classification. For the evaluation of the system, we have used a large dataset, including patients without AMD and patients with non-advanced AMD for evaluation of the proposed system. As it is shown in Figure 4, similar performance as the human observers can be reached using the proposed system. Figure 5 shows example images correctly identified as low risk (5(a)) or as high risk (5(b)) according to both observers.

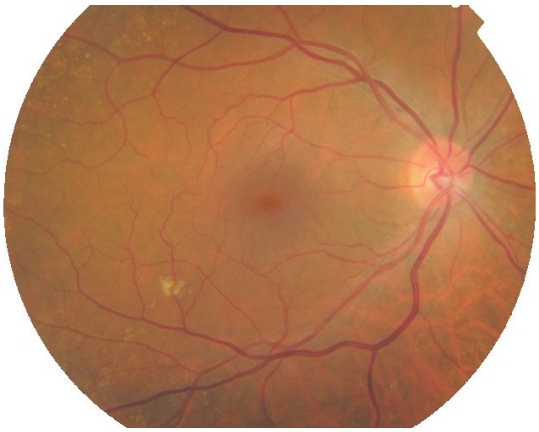


(a) CAD output: low probability to be high risk AMD

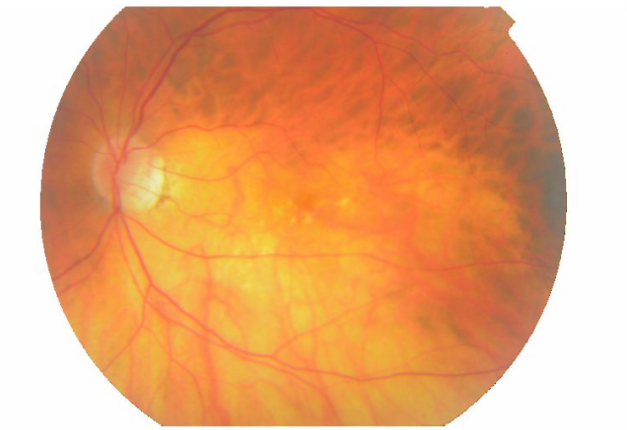
(b) CAD output: high probability to be high risk AMD

Figure 5. (a): Example image where the CAD system outputs a low probability to be high risk and both observers indicated that the image is low risk. (b): Example image where the CAD system outputs a high probability to be high risk and both observers indicated that the image is high risk.

However, the system also has few misclassifications. In Figure 6(a), an example image where the output of the CAD system to be high risk AMD was low, whereas observers indicated that the image was high risk AMD. This could be due to that the image contains only one relatively large drusen within the grid used for grading and the corresponding weighted histogram is not well enough defined in the training set. Adding more examples in the training set might solve this problem. In Figure 6(b), an example image where the CAD output for high risk AMD was high, whereas the observers indicated that the image is low risk AMD. In this image, a relative large number of false positive drusen were detected by the drusen detection algorithm. As a results, the weighted histograms encoding the drusen load accounts for more drusen than the ones actually present in the image. Another reason for misclassifications was bad quality images. Drusen might not be detected due to low contrast or sometimes camera artifacts caused wrong drusen detection. For exact AMD stage identification,



(a) CAD output: low probability to be high risk AMD

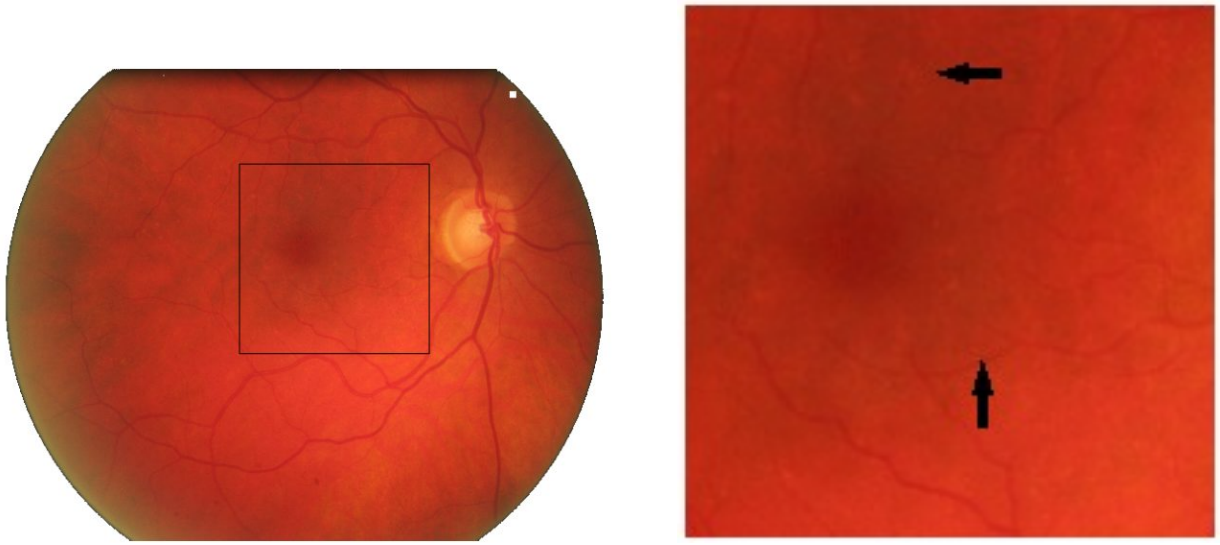


(b) CAD output: high probability to be high risk AMD

Figure 6. (a): Example image where the CAD system outputs a low probability to be high risk, whereas both observers indicated that the image is high risk. (b): Example image where the CAD system outputs a high probability to be high risk, whereas both observers indicated that the image is low risk.

the agreement between the proposed system and the human observers is substantial, but slightly lower than the agreement obtained between observers. The misclassification of the system mainly appeared when distinguishing between no AMD and early AMD. This might be caused because the detection of pigmentary changes, needed to distinguish no AMD from early AMD,[5] is not included in this study. In future work we will investigate the influence of these changes in AMD stage

Another reason might be that some subtle drusen were not detected by the drusen detection system. Therefore, the image is classified into no AMD, while observers indicate that the image is early AMD. Figure 7 shows a color fundus image with very subtle drusen. In this image, not all drusen were detected by the drusen detection system. The image contains very subtle drusen, not picked up by the system. Images classified as early AMD by the system, whereas observers indicated that the image has no signs of AMD mainly correspond to images containing reflections of the internal limiting membrane or containing other non-AMD-related abnormalities. Sometimes, these are also detected as drusen by the drusen detection system, resulting in a weighted histogram containing a false higher drusen load. Improving the drusen detection to remove those false detected drusen could therefore lead to an improvement in AMD stage classification. Other studies



(a) Color fundus image (b) Zoomed in

Figure 7. (a): Example color fundus image containing very subtle drusen. (b) shows the zoomed in region indicated by the black box on (a). Black arrows on (b) indicate locations of very subtle drusen.

have also made an attempt to automatically classify AMD.[9–12] However, the focus of these studies was primarily the identification of images with the presence of AMD. No risk assessment or separation between AMD stages was done in these studies. For clinical use, it is more relevant to know the extension of AMD than only the presence of AMD. This information could be used in a screening setting where patients at high risk could be identified before they actually develop advanced AMD. Doing this manually by human graders would be very costly and time-consuming. An automated system would be a great outcome, reducing the costs substantial for such screening settings.

## 5. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

As far as we are aware, this paper presents for the first time a method to distinguish between high risk and low risk AMD patients and provides an exact AMD stage determination. This is an essential tool for large dataset analysis within population studies and genotype-phenotype correlation analysis. For clinical routine, our method allows for quick and reliable diagnosis of AMD for low cost in a screening settings, allowing for identification of affected, yet asymptomatic patients.

## 6. CONCLUSIONS

The proposed supervised classification method showed even good performance as trained human observers to distinguish low risk AMD patients from high risk AMD patients. Precisely determining the AMD stage is more difficult but experiments showed promising results. The system may be improved by integrating the identification and analysis of pigmentary changes in addition to drusen.

## REFERENCES


[1] Klaver, C., Wolfs, R., Assink, J., van Duijn, C., Hofman, A., and de Jong, P., "Genetic risk of age-related maculopathy. population-based familial aggregation study," *Archives of Ophthalmology* **116**, 1646–51 (1998).

[2] Seddon, J. M., Reynolds, R., Yu, Y., Daly, M. J., and Rosner, B., "Risk models for progression to advanced age-related macular degeneration using demographic, environmental, genetic, and ocular factors," *Ophthalmology* **118**, 2203 – 2211 (2011).

[3] Krishnadev, N., Meleth, A. D., and Chew, E. Y., "Nutritional supplements for age-related macular degeneration," *Current Opinion in Ophthalmology* **21**, 184–189 (2010).

[4] Evans, J. R., "Antioxidant vitamin and mineral supplements for slowing the progression of age-related macular degeneration," *Cochrane database of systematic reviews (Online)* (2006).

[5] Klein, R., Davis, M. D., Magli, Y. L., Segal, P., Klein, B. E. K., and Hubbard, L., "The wisconsin age-related maculopathy grading system," *Ophthalmology* **98**, 1128–1134 (1991).

[6] Scholl, H. P. N., Peto, T., Dandekar, S., Bunce, C., Xing, W., Jenkins, S., and Bird, A. C., "Inter- and intra-observer variability in grading lesions of age-related maculopathy and macular degeneration," *Graefe's Archive for Clinical and Experimental Ophthalmology* **241**, 39–37 (2003).

[7] Quellec, G., Russell, S. R., Seddon, J. M., Reynolds, R., Scheetz, T., Mahajan, V. B., Stone, E. M., and Abràmoff, M. D., "Automated discovery and quantification of image-based complex phenotypes: a twin study of drusen phenotypes in age-related macular degeneration," *Investigative Ophthalmology and Visual Science* **52**, 9195–9206 (2011).

[8] Niemeijer, M., van Ginneken, B., Russel, S. R., Suttorp-Schulten, M. S. A., and Abràmoff, M. D., "Automated Detection and Differentiation of Drusen, Exudates, and Cotton-Wool Spots in Digital Color Fundus Photographs for Diabetic Retinopathy Diagnosis," *Investigative Ophthalmology and Visual Science* **48**, 2260–2267 (2007).

[9] Mora, A. D., Vieira, P. M., Manivannan, A., and Fonseca, J. M., "Automated drusen detection in retinal images using analytical modelling algorithms," *Biomedical Engineering Online* **10** (2011).

[10] Smith, R. T., Sohrab, M. A., Pumariega, N. M., Mathur, K., Haans, R., Blonska, A., Uy, K., Despriet, D., and Klaver, C., "Drusen analysis in a human-machine synergistic framework," *Archives of Ophthalmology* **129**, 40–47 (2011).

[11] Agurto, C., Barriga, E. S., Murray, V., Nemeth, S., Crammer, R., Bauman, W., Zamora, G., Pattichis, M. S., and Soliz, P., "Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images," *Investigative Ophthalmology and Visual Science* **52**, 5862–5871 (2011).

[12] Zheng, Y., Hijazi, M. H. A., and Coenen, F., "Automated "disease / no disease" grading of age-related macular degeneration by an image mining approach." `http://www.iovs.org/content/early/2012/11/12/iovs.12-9576.long` (Accessed November 21, 2012).

[13] van de Ven, J. P. H., Smailhodzic, D., Boon, C. J. F., Fauser, S., Groenewoud, J. M. M., Victor Chong, N., Hoyng, C. B., Jeroen Klevering, B., and den Hollander, A. I., "Association analysis of genetic and environmental risk factors in the cuticular drusen subtype of age-related macular degeneration," *Molecular Vision* **18**, 2271–2278 (2012).

[14] Fauser, S., Smailhodzic, D., Caramoy, A., van de Ven, J. P. H., Kirchhof, B., Hoyng, C. B., Jeroen Klevering, B., Liakopoulos, S., and den Hollander, A. I., "Evaluation of serum lipid concentrations and genetic variants at high-density lipoprotein metabolism loci and TIMP3 in age-related macular degeneration," *Investigative Ophthalmology and Visual Science* **52**, 5525–5528 (2011).

[15] Sánchez, C. I., Niemeijer, M., Isgum, I., Dumitrescu, A. V., Suttorp-Schulten, M. S. A., Abràmoff, M. D., and van Ginneken, B., "Contextual computer-aided detection: Improving bright lesion detection in retinal images and coronary calcification identification in CT scans," *Medical Image Analysis* **16**, 50–62 (2012).

[16] Tax, D. and Duin, R., "Using two-class classifiers for multiclass classification," in [*Pattern Recognition, 2002. Proceedings. 16th International Conference on* ], **2**, 124 – 127 vol.2 (2002).